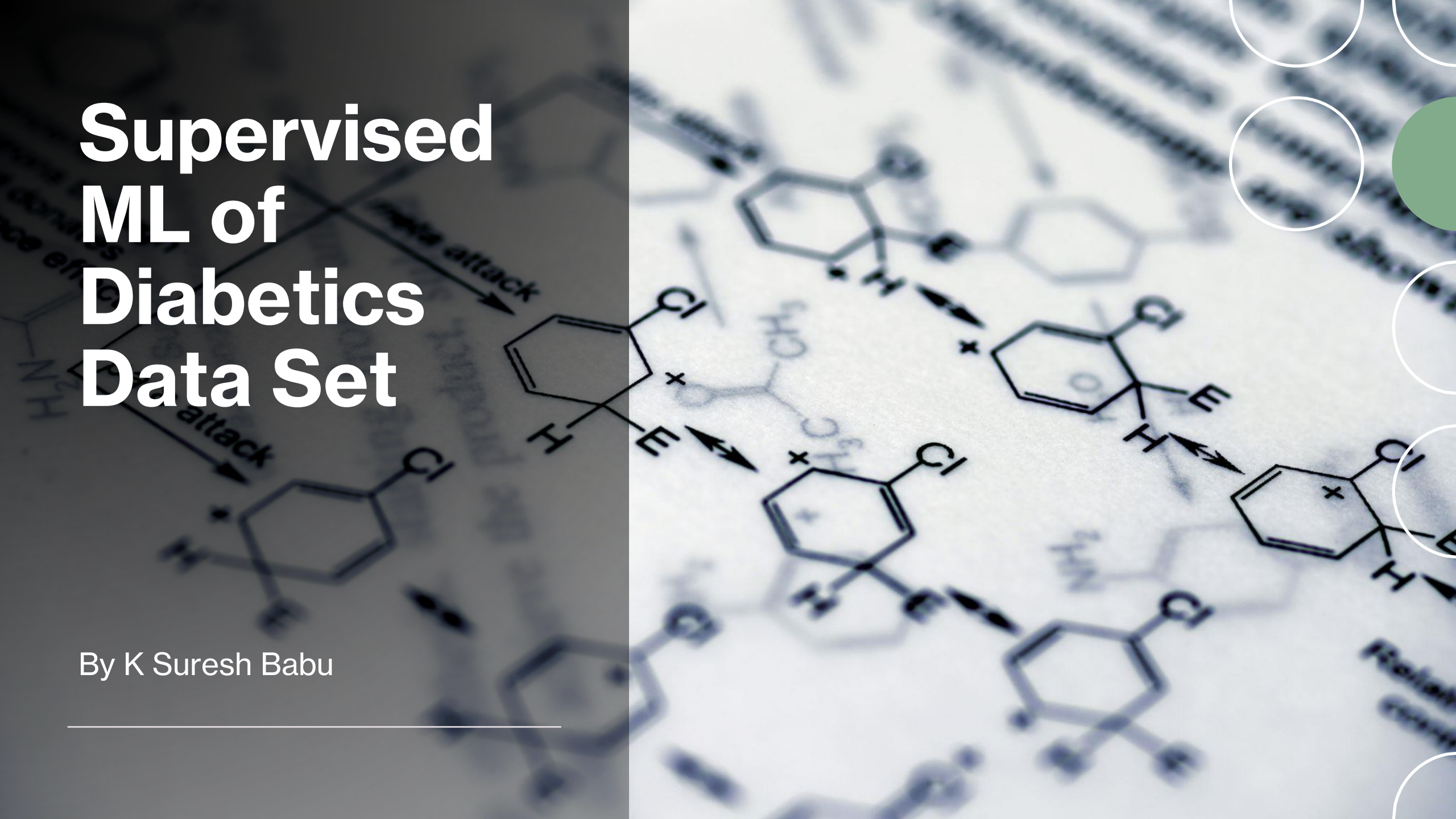# Supervised ML of Diabetics Data Set

By K Suresh Babu

Understanding the Data

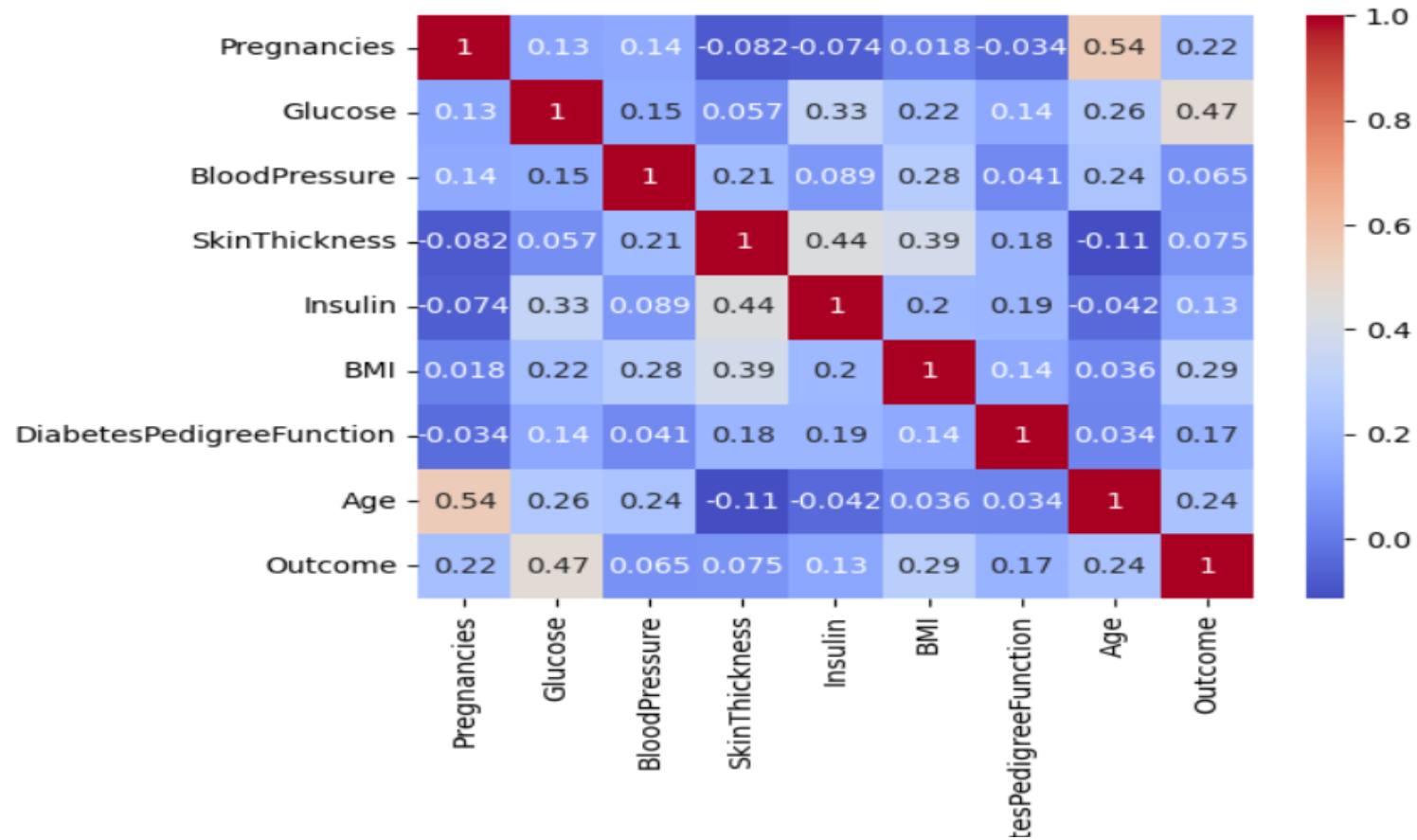The Data set had 768 Patient details with 9 features

The data had good distribution of parameters like age, BMI etc.. for better training.
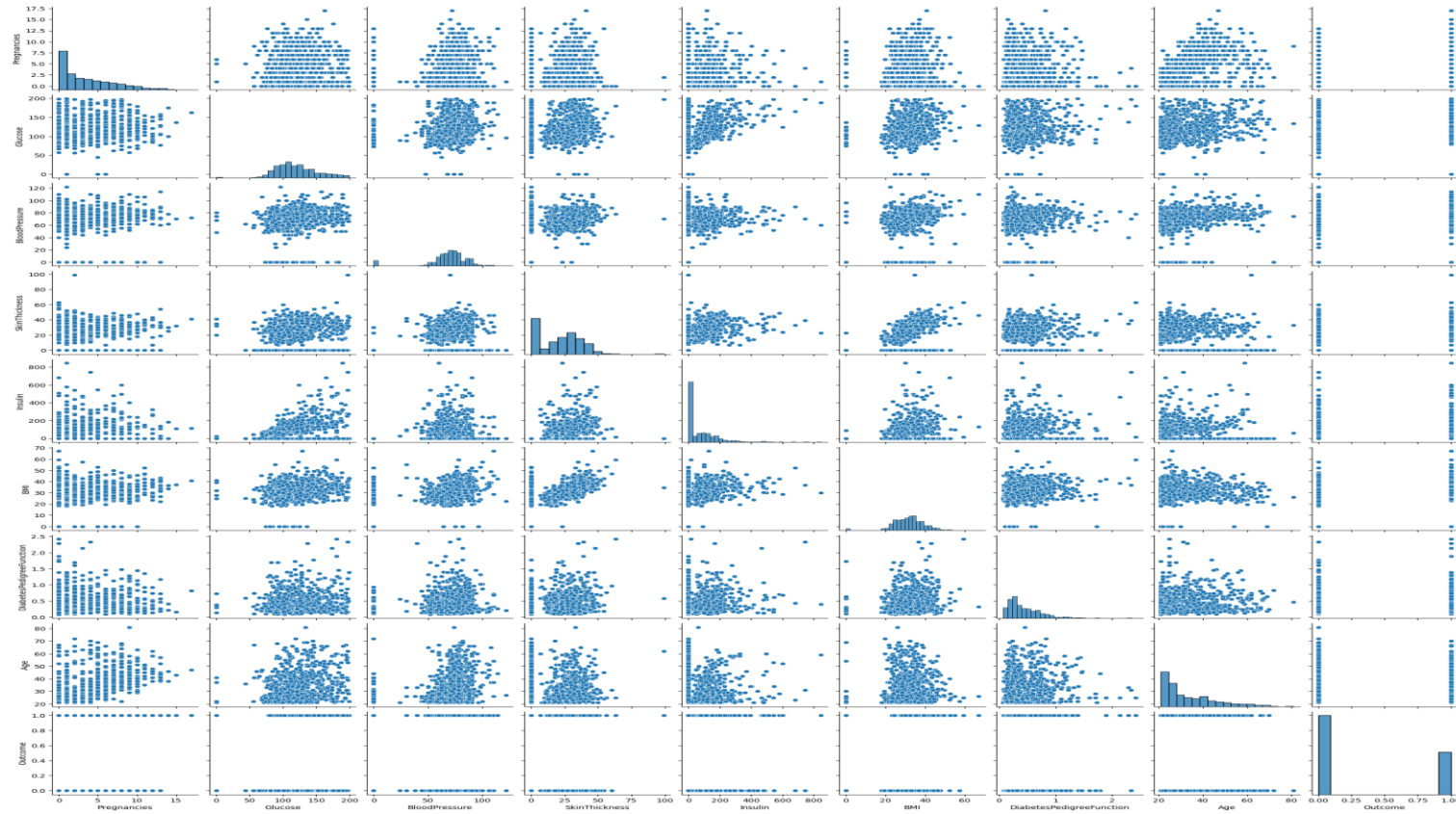
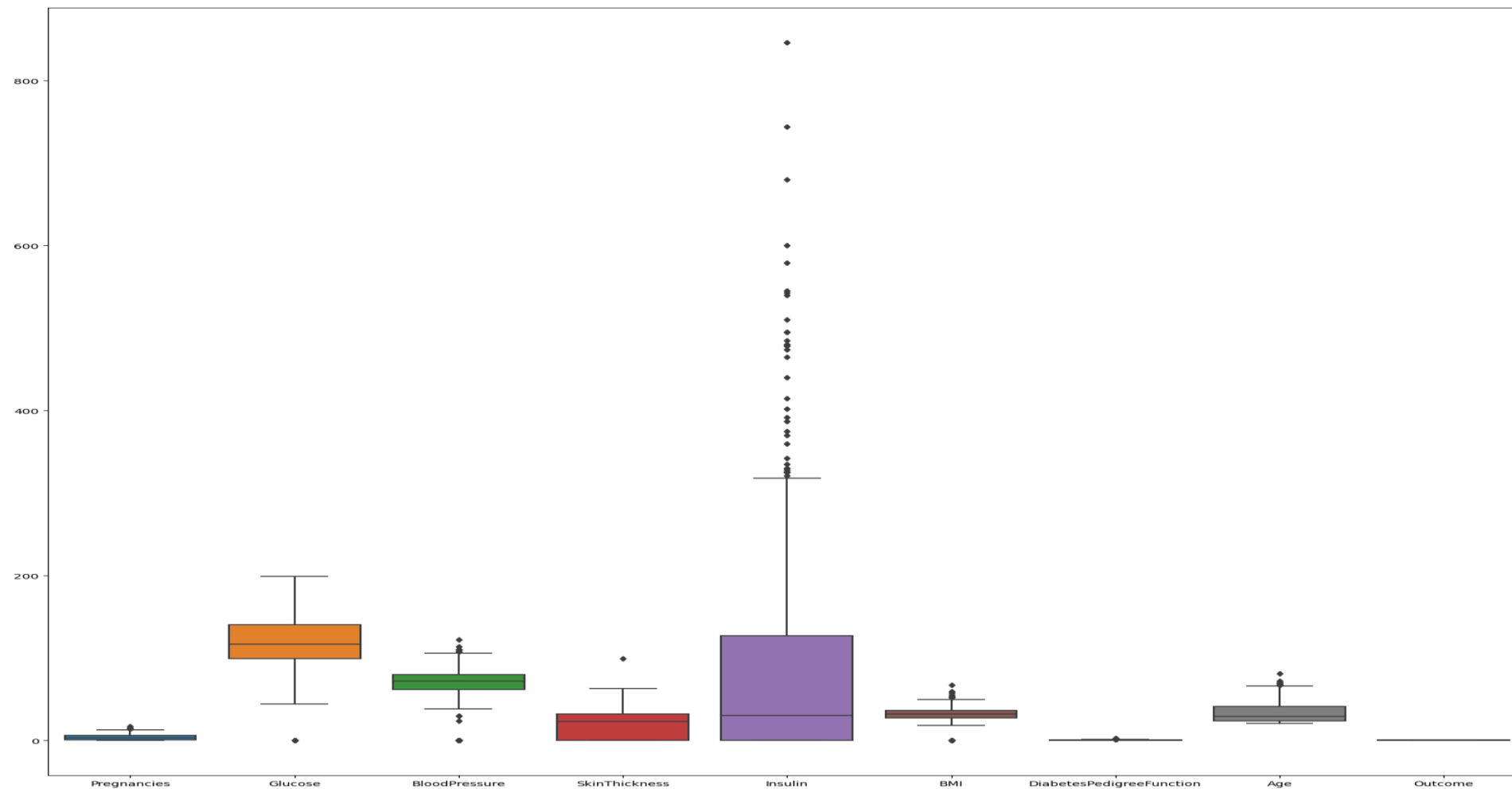The data set had lots of null values and outliers

Correlation with heatmap

EDA of Parameters using Pairplot

Outlier Detection using Box Plot

Preparing the data Set for training a model

- Handling missing values

- Handling outliers

- Scaling and normalization

- Feature Engineering

- Handling imbalanced data

Model Training & Outcome

- Following ML model has been deployed

- 1. Logistics Regression

- 2. KNN - K Nearest Neighbor

- 3. Random Forrest

By comparing the above models -

Logistics regression and Random forest had high score and good prediction capability